BE/APh 161: Physical Biology of the Cell, Winter 2018
Homework #6
Due at the start of lecture, 1PM, February 14, 2018.

Problem 6.1 (Sedimentation, the Einstein-Smoluchowski equation, and the Stokes-Einstein-Sutherland relation, 40 pts)**.**
In this problem, we will derive some landmark results in statistical physics, and learn something about a technique for studying protein structure in the process.

In equilibrium sedimentation experiments, a tube of solution of a protein or protein complex of interest is placed in a centrifuge. The concentration of protein is measured along the tube. The shape of this concentration profile is used to infer information about the size and shape of the protein. For our analysis, let ω be the angular velocity of the rotor of the centrifuge and r describe the distance from the center of the rotor to a given position in the tube of solution. Let $\rho_{\mathrm{H_2O}}$ be the density of the solvent and ρ_{p} be the density of the protein. Note that $\rho_{\mathrm{p}}/\rho_{\mathrm{H_2O}} \approx 1.4$ (BNID 104272). Let a be the radius of gyration of the protein.

a) Due to its density being greater than water, the protein will tend to fall toward the bottom of the tube with steady state velocity v. As it falls through the solvent, it experiences a friction f, such that it experiences a drag force of $F_{\mathrm{drag}} = -fv$. At steady state, the drag force balances the centrifugal force. Use this fact to compute the velocity with which it falls in terms of ρ_{p}, $\rho_{\mathrm{H_2O}}$, a, ω, r, and f. *Hint*: The centrifugal force is given by $F_{\mathrm{centrifugal}} = m_e \omega^2 r$, where m_e is the effective mass of the protein.

b) Show that at steady state, the sedimentation velocity is given by

$$v = D \frac{\mathrm{d} \ln c(r)}{\mathrm{d}r}, \tag{6.1}$$

where D is the diffusion coefficient of the protein.

c) Now use equilibrium statistical mechanics to derive an expression for the concentration profile of the protein. I.e., compute $c(r)$ as a function of $k_B T$ and the other variables describing the system. Note that if $P(r)$ is the probability density for a given particle being at position r in the centrifuge, $c(r) \propto P(r)$. Assume that in the absence of centrifugation, the solution has a uniform concentration of c_0. *Hint*: In part (a), you used an expression for the centrifugal force. Recall that a force $F(r)$ acting on a particle in a potential $U(r)$ is given by $F(r) = -\mathrm{d}U(r)/\mathrm{d}r$.

d) Use your expressions from parts (b) and (c) to derive an expression for D in terms of f. This is the Einstein-Smoluchowski equation, an example of a fluctuation-dissipation theorem. It has this name because it relates equilibrium fluctuations to response to applied perturbations. This is a profound and important concept in statistical physics.

e) The friction f is given by Stokes's law (applicable for spherical particles),

$$f = 6\pi\eta a. \tag{6.2}$$

This was derived by George Stokes by solving for fluid flow around a spherical object. Insert this result into your result in part (d) to get the Stokes-Einstein-Sutherland relation.

f) The sedimentation coefficient S is the ratio of the sedimentation velocity to the acceleration applied to it. It therefore has units of time. Derive an expression for S.

g) Ribosomes are often named by their sedimentation coefficient. A typical unit is a svedberg, which is equal to 10^{-13} s. The 70S ribosome has a sedimentation coefficient of approximately 70 svedbergs. Estimate the diameter of the 70S ribosome. Compare this estimate to what is reported on BioNumbers and explain any discrepancies.

Problem 6.2 (Protein diffusion and temperature, 5 pts)**.**
In what situation might a protein diffuse more *slowly* as the temperature is increased?

Problem 6.3 (Diffusion along a polymer, 5 pts)**.**
Some proteins, such as polymerases, diffuse along DNA prior to finding their binding sites. If a protein diffuses along DNA, its root mean square displacement along the filament scales as $\sqrt{t}$. How does the root mean square displacement *in space* scale with time?

Problem 6.4 (Effects of temperature on pulling polymers, 5 pts)**.**
If I hold the ends of a flexible polymer at a fixed length from each other and then raise the temperature, will it require more or less force to keep the ends at the same distance from each other? Explain.

Problem 6.5 (Polymerase backtracks and random walks, 45 pts)**.**
Sometimes RNA polymerase pauses and then backtracks, pushing the RNA transcript back out the front, as shown in Fig. 1., taken from Depken, et al., *Biophys. J.*, **96**, 2189-2193, 2009.

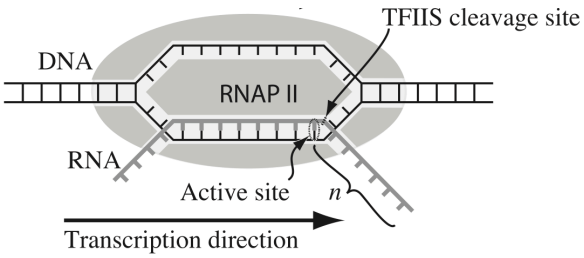

Figure 1: Schematic of RNA polymerase pausing from Depken, et al., *Biophys. J.*, **96**, 2189-2193, 2009

To escape these backtracks, a cleavage enzyme called TFIIS cleaves the bit on RNA hanging out of the front, and the RNAP can then go about its merry way.

Researchers have long debated how these backtracks are governed. Single molecule experiments can provide some much needed insight. The groups of Carlos Bustamante, Steve Block, and Stephan Grill, among others, have investigated the dynamics of RNAP in the absence of TFIIS. They can measure many individual backtracks and get statistics about how long the backtracks last.

One hypothesis is that the backtracks simply consist of diffusive-like motion along the DNA stand. That is to say, the polymerase can move forward or backward along the strand with equal probability once it is paused. This is a one-dimensional random walk. So, if we want to test this hypothesis, we would want to know how much time we should expect the RNAP to be in a backtrack so that we could compare to experiment.

So, we seek the probability density function of backtrack times, $f(t_{\text{bt}})$, where t_{bt} is the time spent in the backtrack. We will approach this problem in two ways. First, we will use a continuum diffusion approach to compute $f(t_{\text{bt}})$. Then, we will simulate the random walks with a computer. In doing the calculations, note that Depken, et al., report that the time it takes to make a base-pair length step in the backtrack random walk is $\tau \approx 0.5$ seconds.

a) First, we take a continuum approach.

   i) In taking a continuum treatment of the backtrack random walk, we consider the diffusion equation,

$$\frac{\partial P(x;t)}{\partial t} = D\,\frac{\partial^2 P(x;t)}{\partial x^2}. \tag{6.3}$$

   What value should you use for the diffusion coefficient D where we take x to be in units of base pairs of the DNA strand?

   ii) Explain why the following choices for initial and boundary conditions are appropriate for modeling the backtrack random walk,

$$P(x;0) = \delta(x+1), \tag{6.4}$$

$$P(0;t) = 0, \tag{6.5}$$

where $\delta(x)$ denotes the Dirac delta function evaluated at x. Note that we define negative x positions of the polymerase as being in a backtrack. We define $P(x \geq 0;t) = 0$, since if x is not negative, the polymerase is not in a backtrack. Recall also that we are taking the values of x to be in units of base pairs.

iii) If you have taken a course in PDEs, solve for $P(x;t)$. Otherwise, demonstrate that

$$P(x;t) = \frac{1}{\sqrt{4\pi Dt}} \left(\mathrm{e}^{-(x+1)^2/4Dt} - \mathrm{e}^{-(x-1)^2/4Dt} \right) \tag{6.6}$$

solves the PDE with the appropriate boundary conditions.

iv) Explain why the probability density function of the amount of time it takes for the polymerase to exit a backtrack, $f(t_{\mathrm{bt}})$, is given by the flux at $x = 0$. That is,

$$f(t_{\mathrm{bt}}) = -D \left. \frac{\partial P(x;t_{\mathrm{bt}})}{\partial x} \right|_{x=0}. \tag{6.7}$$

v) Show that

$$f(t_{\mathrm{bt}}) = \frac{\mathrm{e}^{-1/4Dt_{\mathrm{bt}}}}{\sqrt{4\pi D t_{\mathrm{bt}}^3}}. \tag{6.8}$$

Show that for long t_{bt}, $f(t_{bt}) \sim t_{\mathrm{bt}}^{-3/2}$, a power law behavior.

vi) The cumulative distribution function, $F(t_{\mathrm{bt}})$, or CDF, of the continuous probability distribution is related to the probability density function as

$$F(t_{\mathrm{bt}}) = \int_{-\infty}^{t_{\mathrm{bt}}} \mathrm{d}t'_{\mathrm{bt}} f(t'_{\mathrm{bt}}). \tag{6.9}$$

Show that for long times,

$$F(t_{\mathrm{bt}}) \approx 1 - c t_{\mathrm{bt}}^{-1/2}, \tag{6.10}$$

where c is a positive constant.

b) Now we will approach this problem using a computer simulation. We start at $x = -1$ at time $t = 0$. We "flip a coin," or choose a random number to decide whether we step left or right. We do this again and again, keeping track of how many steps we take and what the x position is. As soon as x becomes nonnegative, we have exited the backtrack. The total time for a backtrack is then τn_{steps}.

i) Write a function that computes the number of steps it takes for a random walker (i.e., polymerase) starting at position $x = -1$ to get to position $x = 0$. It should return the number of steps to take the walk.

ii) Generate 10,000 of these backtracks in order to get enough samples out of $f(t_{\text{bt}})$. The calculation may take a few minutes. If you are using Python and are interested in a way to really speed up this calculation, check out Numba.

iii) It is often useful to display probability distributions using their **empirical cumulative distribution function**, or ECDF. If you have many samples out of a distribution (you now have 10,000 samples out of $P(t_{\text{bt}})$), the ECDF is defined as

$$\text{ECDF}(x) = \text{fraction of samples } \leq x. \tag{6.11}$$

In this case, it is more informative to plot the **empirical complementary cumulative distribution function** ECCDF, which is

$$\text{ECCDF}(x) = 1 - \text{ECDF}(x). \tag{6.12}$$

Plot the ECCDF of your samples. Add a plot of the CCDF, which is given by $1 - F(t_{\text{bt}})$. You need only plot the large t limit of the CCDF. Describe how this shows the power law nature of the probability density function $f(t_{\text{bt}})$. *Hint*: Think about what axes you should plot on a log scale to make this plot informative.

c) Interestingly, many researchers thought (and maybe still do) there were two classes of backtracks: long and short. Is the hypothesis that the backtrack is a random walk process commensurate with seeing both very long and very short backtracks? Explain your reasoning.